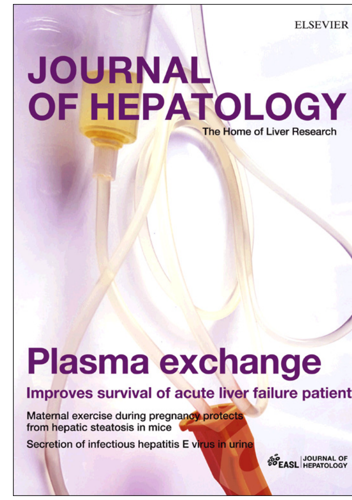
DOI:
10.1016/j.jhep.2017.03.014

Document Version
Peer reviewed version

Citation for published version (APA):
Welzel, T., Bhardwaj, N., Hedskog, C., Chodavarapu, K., Camus, G., McNally, J., Brainard, D., Miller, M. D., Mo, H., Svarovskaia, E., Jacobson, I., Zeuzem, S., & Agarwal, K. (2017). Global Epidemiology of HCV Subtypes and Resistance-associated Substitutions Evaluated by Sequencing-Based Subtype Analyses. *Journal of Hepatology*. https://doi.org/10.1016/j.jhep.2017.03.014



Global Epidemiology of HCV Subtypes and Resistance-associated Substitutions Evaluated by Sequencing-Based Subtype Analyses

Tania Welzel, Neeru Bhardwaj, Charlotte Hedskog, Krishna Chodavarapu, Gregory Camus, John McNally, Diana Brainard, Michael D. Miller, Hongmei Mo, Evguenia Svarovskaia, Ira Jacobson, Stefan Zeuzem, Kosh Agarwal

PII: S0168-8278(17)30148-4
DOI: http://dx.doi.org/10.1016/j.jhep.2017.03.014
Reference: JHEPAT 6465

To appear in: *Journal of Hepatology*

Received Date: 5 August 2016
Revised Date: 10 February 2017
Accepted Date: 6 March 2017

Title: Global Epidemiology of HCV Subtypes and Resistance-associated Substitutions Evaluated by Sequencing-Based Subtype Analyses

Authors: Tania Welzel,[1] Neeru Bhardwaj,[2a] Charlotte Hedskog,[2] Krishna Chodavarapu,[2] Gregory Camus[2], John McNally,[2] Diana Brainard,[2] Michael D. Miller,[2] Hongmei Mo,[2] Evguenia Svarovskaia,[2] Ira Jacobson,[3] Stefan Zeuzem,[1] Kosh Agarwal.[4]

[1] Johann Wolfgang Goethe University Medical Center, Frankfurt am Main, Germany

[2] Gilead Sciences, Foster City, CA, USA

[3] Mount Sinai Beth Israel Medical Center, New York, USA

[4] King's College Hospital Foundation Trust, London, UK

[a] Corresponding author: Neeru Bhardwaj, 333 Lakeside Drive Foster City, CA, USA 94404; Phone: +1 (650)-235-3286; Neeru.Bhardwaj@gilead.com

Keywords (list up to 12): Hepatitis C virus, subtypes, resistance, deep sequencing.

Word Count (Introduction through References): 6,334

Number of Figures: 3

Number of Tables: 5

List of Abbreviations:

aLRT: Approximate likelihood-ratio test

DAA: Direct-acting antiviral

GT: Genotype

HCV: Hepatitis C virus

NS: Non-structural

RAS: Resistance-associated substitution

RdRp: RNA-dependent RNA polymerase

SOF: Sofosbuvir

US: United States

UTR: Untranslated region

Financial Support:

Financial support was provided by Gilead Sciences, Inc.

Conflict of Interest: Drs. Bhardwaj, Hedskog, Camus, McNally, Brainard, Miller, Mo and Svarovskaia are employed by and own stock in Gilead Sciences, Inc. Dr. Chodavarapu owns stock in Gilead Sciences, Inc. Dr. Welzel is a consultant and speaker for Abbvie, Boehringer-Ingelheim, Bristol-Myers Squibb, Janssen and Gilead Sciences, Inc. Dr. Zeuzem is a consultant for Abbvie, Bristol-Myers Squibb, Janssen, Merck/MSD and Gilead Sciences, Inc. Dr. Jacobson is a speaker and/or consultant for for Abbvie, Achillion, Bistrol Myers Squibb, Enanta, Janssen, Merck, Tobira and Gilead Sciences, Inc. Dr. Agarwal is a speaker and/or consultant for AbbVie, Achillion, Astellas,

Bristol-Myers Squibb, GlaxoSmithKline, Janssen, Merck, Novartis and Gilead Sciences, Inc.

Author Contributions: Concept and Design – TW, NB, CH, KC, GC, JM, DB, MM, HM, ES, IJ, SZ, KA. Experiments and Procedures – TW, NB, CH, KC, GC, JM, DB, MM, HM, ES, IJ, SZ, KA.. Writing and Review – TW, NB, CH, KC, GC, JM, DB, MM, HM, ES, IJ, SZ, KA.

Abstract:

Background & Aims: Hepatitis C virus (HCV) is highly variable with 7 genotypes and 67 subtypes characterized to date. HCV genotype (GT) and subtype, and presence of resistance associated substitutions (RASs) represent the key viral determinants for the selection of direct-acting antiviral (DAA) treatment regimens. However, currently available HCV genotyping assays have limitations in differentiating between HCV subtypes and RAS prevalence in different subtypes is largely undefined.

Methods: In this study, we compared HCV GT and subtypes with INNO-LiPA 2.0 vs. amplicon sequencing among 8,945 patients from phase II/III clinical trials of DAAs. We also investigated global HCV molecular epidemiology in 12,615 patients. Subtype RAS prevalence was determined by population or deep sequencing, and phylogenetic analyses investigating subtype diversity were performed.

Results: Though there was high concordance between INNO-LiPA and sequencing for GT determination, INNO-LiPA was insufficient for subtype determination for GT2, 3, 4, and 6. Sequencing provided subtype refinement for 42%, 10%, 81%, and 78% of GT2, 3, 4, and 6 patients, respectively. GT discordance (GT2-GT1) was observed in 28 of 950 (3%) GT2 patients, consistent with inter-genotype recombinants. Sequencing-based analyses demonstrated variations in regional subtype prevalence, notably within GT2, 4 and 6. RAS prevalence varied by subtype, with the clinically relevant NS3 RAS Q80K found in GT1a, 5a and 6a and the NS5A RAS Y93H in GT1b, 3a, 4b, 4r and 7.

Conclusions: Together, these analyses provide an understanding of subtyping accuracy and RAS distribution that are crucial for the implementation of global HCV treatment strategies.

Lay Summary:

The sequencing analyses performed on 12,615 patients with HCV infection investigated three main questions: (1) the concordance between NS5B amplicon sequencing and INNO-LiPA for HCV genotype (GT) determination; (2) the prevalence of HCV subtypes for each genotype around the world; and (3) the prevalence of resistance-associated substitutions (RASs) in different subtypes. Overall, there was high INNO-LiPA and sequencing concordance for GT determination, but subtype determination for GT2, 3, 4, and 6 needed sequencing refinements. Sequencing-based analyses demonstrated variations in regional subtype prevalence for some genotypes, notably within GT2, 4 and 6, and RAS prevalence varied by subtype, with the clinically relevant NS3 RAS Q80K found in GT1a, 5a and 6a and the NS5A RAS Y93H in GT1b, 3a, 4b, 4r and 7.

Introduction:

Hepatitis C virus (HCV) is a major human liver pathogen. It is estimated that >170 million people are affected worldwide and at risk for progressive liver disease and hepatocellular carcinoma (1, 2). HCV is classified into 7 genotypes and 67 subtypes, and displays ~30% divergence at the genotype level and ~15% divergence at the subtype level (3). HCV diversity is derived in part from its RNA-dependent RNA polymerase (RdRp) that lacks proofreading capability, recombination, genetic drift, and immune pressures on the virus (4). The treatment of HCV infection was revolutionized with the advent of highly effective direct-acting antiviral (DAA) treatment regimens that target HCV gene products, such as the NS5B RdRp, the NS5A phosphoprotein and NS3 protease (5).

Similar to treatment with pegylated-interferon and ribavirin, selection of a DAA regimen, treatment duration and sustained virologic response (SVR) rate are dependent on HCV genotype (GT) and subtype (6-9). Previous investigations have shown that HCV GT distribution varies worldwide (10, 11), though less is known about HCV subtypes. HCV subtype sequences have been shown to cluster on phylogenetic trees based on region of origin (12-14). Potentially, the diversification of HCV in local epidemics could affect SVR rates if resistance associated substitutions (RASs) emerge.

The prevalence of recombinant forms of HCV, detected in Eastern Europe and Russia (15-17), are not known on a global scale. Identification of recombinant sequences will inform the global prevalence of known recombinant viruses, like RF2k/1b, and potentially identify new recombinant forms in the population. It is important to understand the prevalence of HCV endemic in different parts of the world as it could optimize the selection of therapeutic regimens and treatment durations.

The presence of RASs for DAAs targeting NS3, NS5A and NS5B has not been systematically assessed across all HCV subtypes (18, 19). RASs vary at the subtype level (20) and could affect treatment outcome when present at baseline. Thus, in the absence of sequencing for each patient, the assessment of RASs in the context of HCV subtype should be considered for optimal DAA therapeutic regimens.

The aim of this study was to investigate HCV epidemiology in 12,615 patient samples from 28 different countries across five geographic regions. As the accuracy of current genotyping methods may vary, we first compared subtype classification by INNO-LiPA, the current standard (21, 22), to sequencing of NS3/4A, NS5A and NS5B amplicons in a subset of 8,945 patient samples. We next investigated prevalence of HCV subtypes, and subtype-specific RASs across the different geographic regions.

Materials and Methods:

Clinical samples

The 12,615 patients analyzed provided written informed consent for enrollment in 67 clinical studies run by Gilead Sciences (Foster City, CA) and Pharmasset (formerly Princeton, NJ). The studies and numbers of patients are outlined in Supplementary Table 1. Study protocols followed the ethical guidelines set in place by the 1975 Declaration of Helsinki and were approved by the relevant institutional review board committees. All samples included in this analysis are baseline samples collected from treatment naïve and experienced patients. More than 95% of patients included were DAA-naïve.

Genotyping using INNO-LiPA and amplicon sequencing

The Siemens VERSANT HCV Genotype INNO-LiPA 2.0 Assay (Innogenetics, Ghent, Belgium) was performed by Covance (Indianapolis, IN; Singapore; Geneva, Switzerland), QLabs (Livingston, UK; Marietta, GA; Singapore) and Cenetron (Austin, TX). The genotype assignment from INNO-LiPA was used to select genotype-specific primers located outside of the gene target(s) that amplify the entire NS3/4A, NS5A, or NS5B regions in patient samples. Depending on the clinical trial, some patients had multiple gene targets amplified. Standard reverse transcription polymerase chain reaction (RT-PCR) was performed on patient plasma with HCV RNA >1000 IU/mL at DDL Diagnostic Laboratory (Rijswijk, The Netherlands). If amplification using genotype specific primers failed, a partial NS5B sequence was amplified using genotype independent primers, allowing for sequence-based genotype/subtype classification of the sample.

For deep sequencing, amplicons encoding the subject-derived NS3/4A, NS5A and NS5B were run using Illumina MiSeq v2 150 paired-end deep sequencing at DDL or WuXi AppTec (Shanghai, China). FASTQ files were split based on 100% matched barcodes. Contigs were generated from paired-end FASTQ files using VICUNA (Broad Institute, Cambridge, MA) and merged to create a de novo assembly sequence. All paired-end reads were merged using PEAR (Exelixis Lab, Heidelberg, Germany), chopped at the 3' end when MAPQ<15, and filtered to remove reads <50 bases. The filtered reads were aligned to the de novo assembly sequence using MOSAIK v1.1.0017 (Marth Lab, Boston, MA) to create a final assembly sequence. The aligned reads were translated in-frame and the resulting tabulated summary of variants from the final assembly was utilized to generate a consensus sequence. Mixtures were reported when present ≥15% of the viral population. Population sequences captured variants at 15% prevalence.

NS3/4A, NS5A and NS5B consensus nucleotide and amino acid sequences were compared by the NCBI alignment tool BLAST (23) to a set of reference sequences to assign HCV genotype and subtype. A total of 170 NS3, 167 NS5A, and 169 NS5B reference sequences representing 7 genotypes and 67 subtypes were used for comparison. Each subtype was represented by 1-7 sequences in this reference set. A definitive genotype and subtype was indicated when its nucleotide and amino acid sequence had ≥85% homology to a reference sequence. The 85% threshold for the assignment of a subtype was designed to be conservative to allow subtypes to be assigned and to prevent spurious assignments. The subtype with highest homology to a sequence at the nucleotide and amino acid level was reported for each sample. If multiple targets were sequenced, the subtype analysis results in multiple subtypes assigned to a patient sequence. In Table 1, a sample was labeled as “novel/mixed” if the closest reference sequence did not show at least ≥85% homology, if the genotype and subtype defined by nucleotide and amino acid homology were not identical, or if the subtype assignments were not identical between amplicons. In Table 2 and Figures 1, only NS5B was used to determine subtype. A sample was labeled “novel/mixed” if there was no subtype match with at least 85% homology, or if the top nucleotide and amino acid subtype hits was not identical. A sample was reported to be a GT2-GT1 recombinant if the INNO-LiPA result indicated GT2, while NS3/4A, NS5A and/or NS5B sequencing indicated GT1 and full genome sequencing verified a recombination breakpoint in the virus (24).

Concordance of INNO-LiPA and sequencing

The genotype and subtype defined by INNO-LiPA and sequencing analyses were compared at baseline for 8,945 patients in Table 1, and categorized as concordant or discordant at the genotype level. At the subtype level, results were concordant when the subtype reported by INNO-LiPA matched sequencing, as in a result of GT2a by INNO-LiPA and GT2a by sequencing. A result was categorized as refined when the correct genotype or subtype group was reported by INNO-LiPA but was not exact, as in a result of GT2 or GT2a/2c by INNO-LiPA and GT2a by sequencing. A result was discordant when an incorrect genotype or subtype was reported by INNO-LiPA as compared to sequencing. A result of novel/mixed was reported when sequencing could not define the exact sample subtype.

Global Prevalence of HCV subtypes

The prevalence of HCV subtypes was estimated by using the HCV subtype determined by baseline NS5B sequencing for 12,615 patients and their associated country of origin in Table 2 and Figure 1. A virus was categorized as novel/mixed if the NS5B nucleotide and amino acid reference sequences determined by BLAST were not concordant. Subtype prevalence was calculated as: # of subtype in country/# of genotype in country. The region (and countries) reported are: North America (US, Canada and Puerto Rico), Europe (Great Britain, Ireland, Sweden, Germany, Netherlands, Belgium, France, Spain, Italy, Switzerland, Poland and Estonia), Asia (Russia, India, China, Hong Kong, Taiwan, Korea, Japan and Thailand), Oceania (Australia and New Zealand), and Africa (Egypt and South Africa). NS5B sequences representing genotypes 1c, 1e, 1g, 1h, 1l, 2i, 2j, 2k, 3g, 3i, 4c, 4f, 4k, 4l, 4n, 4o, 4r, 4t, 6e, 6f, 6h, 6i, 6l, 6m, 6n, 6p, and 6q were deposited into GenBank (KY608611-KY608718).

RAS Prevalence by HCV subtype

The subtype prevalence of HCV RASs was determined for NS3 (n=2,664), NS5A (n=7,893) and NS5B (n=11,074) in patients at baseline. Patients with DAA-experience were excluded from analysis. For each subtype, patient amplicon amino acid sequences were aligned and RAS positions analyzed for frequencies of amino acids present at the noted location. RASs were detected with ≥15% prevalence in a patient. Of the RASs detected, those with ≥5% prevalence within the subtype population were reported in Tables 3-5. The majority and minority variants for each subtype were reported relative to the GT1a H77 reference amino acid at the indicated position.

Phylogenetic Analysis

Individual phylogenetic analyses were performed for NS5A and NS5B sequences from patients with GT3a and GT1b infection. Sequences were aligned using ClustalW (25). The NS5B GT3a phylogenetic tree included 351 sequences (Australia n=25, Canada n=47, Germany n=16, Spain n=13, Estonia n=3, France n=12, England n=41, India n=31, Italy n=12, Netherlands n=2, New Zealand n=19, Poland n=2, Puerto Rico n=2, Russia n=45, Sweden n=3, US n=78). The NS5B GT1b phylogenetic tree included 621 sequences (Australia n=8, Belgium n=8, Canada n=3, England n=8, France n=17, Germany n=62, India n=47, Italy n=48, Ireland n=2, Japan n=70, Korea n=86, New Zealand n=39, Poland n=39, Puerto Rico n=7, Spain n=34, Taiwan n=67, USA n=76). Maximum likelihood phylogenetic trees were inferred using default parameters in GARLi (version 2.0), which optimizes the substitution model iteratively (26). The statistical significance of the branches was assessed by approximate likelihood-ratio test (aLRT) (27) using PhyML 3.0 (28) and a consensus tree was generated in PAUP (version 4.0b10). Branches with p-values of <0.001 were considered supported. The phylogenetic trees were visualized using FigTree (version 1.3.1). A cluster is defined as phylogenetically linked sequences which are separated from all other sequences by a significant branch.

Results:

Comparison of genotype and subtype by INNO-LiPA and amplicon sequencing

INNO-LiPA and amplicon sequencing were performed for 8,945 HCV-infected patients in order to provide HCV genotype (GT) and subtype. Based on INNO-LiPA results, 5,701 patients were GT1, 950 were GT2, 1,841 were GT3, 309 were GT4, 37 were GT5 and 107 were GT6. There was a high level of concordance between these two approaches, as shown in Table 1. Greater than 99% agreement between INNO-LiPA and amplicon sequencing was observed for GT1 (99.8%), GT3 (99.9%), GT4 (99.7%), GT5 (100%) and GT6 (100%), whereas GT2 showed a concordance rate of 96.9%. As a whole, 8,904/8,945 (99.5%) genotypes were concordant between INNO-LiPA and amplicon sequencing.

The GT2 concordance rate was ~3% lower than all other assayed genotypes. 29 patient samples were presumably incorrectly typed by INNO-LiPA as GT2. 28/29 of these GT2 discordant samples were typed as GT1 by amplicon sequencing. 17/29 of these genotype discordant samples were investigated using full genome sequencing. In 16/17 of these cases, the virus was an HCV recombinant between GT2 (5'UTR to NS2) and GT1 (NS3-3'UTR), with breakpoints present at the NS2-NS3 border, as previously described (29). 6 of these 16 GT2-GT1 recombinants were identified as the circulating RF2k/1b strain and were detected in patients from Belgium (2/6), the Netherlands (2/6), Spain (1/6), and the US (1/6). Another 6 of the 16 characterized GT2-GT1 recombinants were GT2b/1a recombinants and all were found in the US. The remaining 4 GT2-GT1 recombinants were GT2b/1b recombinants found in the US (3/4) and Puerto Rico (1/4). 12 of these 16 GT2-GT1 recombinants were characterized previously (29). Of interest, the remaining characterized discordant GT2 sample (1/17) was identified to be GT7 (30).

Genotype discordant samples classified as GT1, 3 and 4 by INNO-LiPA were also investigated. Of the 10 patient samples found to be GT1 discordant, 8/10 where classified as GT6 by amplicon sequencing. Full genome sequencing analysis performed on 6/8 of these GT1 discordant samples corroborated that the viruses were GT6. There were only 2 patient samples that were either GT3 (n=1) or GT4 (n=1) discordant. In both cases, the virus was identified to be GT1 by amplicon sequencing. Similar to amplicon sequencing, the full genome sequencing analysis completed on the discordant GT3 sample also defined the sample genotype as GT1.

The concordance rate between INNO-LiPA and amplicon sequencing at the subtype level was more variable than at the genotype level. Subtyping for GT5 and GT1 were the most successful, with 100% and 97.8% of the assigned subtypes corroborated by amplicon sequencing. The remaining genotypes had lower levels of subtype concordance. INNO-LiPA subtyping for GT2, 3, 4 and 6 had concordance rates of 51.1%, 90.1%, 5.8% and 9.3% with amplicon sequencing and required refinement in 41.5%, 9.5%, 81.2% and 77.6% of the samples, respectively. In these genotypes, one of the causes for INNO-LiPA refinement is due to the categorization of subtypes into groups, as in the subtype groups 2a/2c, 4a/4c/4d and 6c-l. Subtype discordance, where the subtype assignment was not captured by INNO-LiPA, was low in GT1 (0.6%), 3 (0.3%) and 5 (0%). In contrast, GT2, 4 and 6 had higher discordance levels of 4.4%, 10.7% and 8.4%, respectively. Furthermore, novel and/or mixed subtypes were observed at higher rates in GT2 (3.1%), 4 (2.3%) and 6 (4.7%).

Geographic distribution of HCV subtypes

HCV subtype distribution for each genotype was analyzed in a data set of 12,615 patients representing 7 HCV genotypes, 28 countries and 5 geographic regions, namely North America (n=7,891), Europe (n=2,166), Asia (n=977), Oceania (n=1,474) and Africa (n=107). NS5B sequencing was used to assign subtypes to all patients. 46 previously characterized HCV subtypes from GT1-6 were captured in this analysis, identifying 7 GT1, 7 GT2, 6 GT3, 12 GT4, 1 GT5, and 13 GT6 subtypes. GT1-6 had a novel/mixed subtype category included due to the large number of sequences that did not match characterized strains. In addition, GT2-GT1 recombinant sequences and a genotype 7 sequence were surveyed. Thus, in total, 54 subtypes of HCV and their global distribution are shown by region in Table 2 and by country in Figure 1 and Supplementary Table 2.

In our sampling of GT1 (n=7,906), subtypes 1a and 1b were most prevalent across all countries. North America and Oceania was 76.4-78% subtype 1a and 21.8-23.2% 1b, whereas Europe was 49.9% 1a and 48.4% 1b, and Asia was 95.6% 1b and 4.1% 1a (Table 2). Interestingly, Poland, though included as a part of Europe, followed the Asian distribution and was 100% 1b. Novel/mixed GT1 sequences were detected in Great Britain (1.3%), France (0.8%) and India (4%), as well as in the US (0.1%) at a lower frequency. HCV GT2-GT1 recombinant strains (RCF2k-1b, GT2b-1a and GT2b-1b), included under GT1, make up 0.2% of sampled strains in North America and 0.5% in Europe.

Unlike GT1, the circulating subtypes of GT2 (n=1,408) vary widely. North America is predominantly subtype 2b (82.7%) and 2a (15.6%), while Asia is predominantly subtype 2a (95.9%) followed by 2b (4.1%). Oceania is split between the three subtypes 2b (59.6%), 2a (35.4%) and 2c (5.1%). Interestingly, in Europe, there is no consensus in circulating GT2. Germany and Great Britain appear to mirror Oceania in being primarily subtypes 2b (55-63.3%) followed by 2a (13.3-25%) and 2c (10-15%), with the notable exception that 2i, 2j and 2k subtypes are detectable in the population. Italy has a unique GT2 subtype profile, with 100% 2c, a subtype which only appeared <34% in other locations. Most remarkably, a high proportion of circulating GT2 subtypes were novel/mixed in the Netherlands (60%), Belgium (33.3%) and France (28.3%). In addition, these countries had circulating subtypes 2a, 2b, 2c, 2i and 2k, giving them a divergent profile from other regions that were primarily 2a, 2b and/or 2c.

For GT3 (n=2,570), subtype 3a is most prevalent in North America (98.7%), Europe (98.9%) and Oceania (98.7%). However, the subtype breakdown in Asia is distinctive and is driven by the GT3 subtypes circulating in India. 55.9% of the individuals sampled in India were infected with subtype 3a, however subtype 3b (20.3%), 3g (6.8%), 3i (13.6%) and uncharacterized novel/mixed (3.4%) were present in the population as well. By contrast, Russia, the other country included as part of Asia, was 100% subtype 3a.

Similar to GT2, GT4 (n=526) subtype distribution varied widely, particularly in Europe. In our sampling, subtype 4a dominates North America (82.4%). Subtypes 4a (45.5%), 4d (45.5%) and 4n (9.1%) were observed to be circulating in Oceania. Africa, which in this study is represented by samples collected in Egypt, was primarily subtype 4a (84.3%), followed by novel/mixed (9.8%), 4o (4.9%) and 4l (1%). It is notable that Egypt has novel/mixed GT4 subtypes as HCV spread in the population in the 1960-80s during an anti-schistosomiasis campaign (31), presumably causing clonal infections. In Europe, the prevalence of GT4 subtypes changes per country, but is predominantly 4a (42.2%) and 4d (33.9%), similar to Oceania. However, Spain was the only country assayed to have circulating subtype 4b (72.7%). Belgium (23.1%), Germany (9.1%), Great Britain (5.6%) and France (3.4%) have high levels of circulating novel/mixed strains. Of note, GT2 novel/mixed strains were also observed in Belgium, Great Britain and France.

In GT5 (n=83), subtype 5a is the prevailing subtype in every tested country. The only exception to this observation is one patient in France who had a novel/mixed GT5 strain detected. Even with a restricted sample population in GT6 (n=121), GT6 subtypes showed great diversity in all regions tested. The only characterized subtype to be sampled in North America (37.5%), Europe (33.3%), Asia (74.3%) and Oceania (33.3%) was 6a, but was present at varying frequency. Another commonality between regions was the detection of novel/mixed strains that occurred in the US (7.4%), France (33.3%), and New Zealand (8.3%).

HCV subtype distribution of resistance-associated substitutions

In our analysis of sequence data, a cutoff of 15% was used to call the presence of a polymorphism in a patient's HCV consensus sequence. In order to determine the prevalence of RASs in different subtypes, we reported the presence of a RAS within a subtype when at least 5% of the patients sampled had the RAS present in their consensus sequence. NS3 (n=2,664), NS5A (n=7,893) and NS5B (n=11,074) sequences were analyzed for this RAS analysis. Baseline RAS frequencies in different

HCV subtypes are summarized for NS3 in Table 3. The NS3 RAS position Q80K has been shown to have high prevalence in GT1a as compared to GT1b (20) and testing for baseline Q80K is recommended prior to treatment with the NS3/4A inhibitor simeprevir due to reduced SVR rates (32). Our analysis corroborates a high baseline prevalence of Q80K in GT1a (36%), but not in GT1b (1.7%). Q80K was also detected in GT5a (100%) and GT6a (100%). 100% of HCV GT2 subtypes carried Q80G, however this is not considered a resistance-conferring substitution. V36L was found in the majority of all GT2, 3, 4 and 5 subtypes and in the 1 GT7 sequence assayed in our data set. Though not sampled extensively, GT1c (n=1; R155D, A156T, D168E), 1e (n=2; V36L, T54S, S122N) and 7 (n=1; V36L, V55A, D168Q) showed at least three NS3 RASs at baseline. Other major NS3 RASs found in HCV subtypes were D168E in GT5a (53%) and 6a (7%).

Baseline NS5A RAS frequencies are shown by HCV subtype in Table 4. As compared to NS3, there was much higher diversity at RAS positions in NS5A. The NS5A RAS Y93H is associated with reduced NS5A DAA efficacy, with or without L31M/V/I in GT1b patients (33). Baseline Y93H was identified in GT1b (11%), 3a (6%), 4b (50%), 4r (13%), and in the 1 GT7 patient. Other substitutions assayed at ≥50% at this position were Y93F (GT1g), Y93S (GT6f, 6m, 6n) and Y93T (GT6a, 6e, 6f, 6g, 6h, 6l, 6o, 6p, 6q, 6r). The related NS5A RAS L31M/V/I, which increases resistance conferred by Y93H to NS5A inhibitors (34), was detected in the Y93H carrying subtypes GT1b (4%), 4b (100%) and 4r (25%), though not in GT3a or 7. L31M/V/I was sampled frequently across many GT2, 3 and 4 subtypes. Of note, all GT4r patients carried NS5A RASs at baseline. 3 of 8 GT4r patients carried triple NS5A RASs

(M28M/V+Q30R+L31M or M28M/V+Q30R+Y93H), 4 of 8 carried double NS5A RASs (M28M/V/I+Q30R) and 1 of 8 GT4r patients a single NS5A RAS (Q30R) at baseline.

Baseline NS5B RASs were rare and are shown by HCV subtype in Table 5. The S282T mutation, which confers resistance against the NS5B inhibitor sofosbuvir (SOF) (35), was not found at any appreciable frequency in the assayed subtypes.

Intra-subtype diversity

Viral diversity within HCV is represented by a high number of characterized genotypes and subtypes, as is seen with GT4 and 6, but also by the variation within subtypes. Both GT1b and 3a are subtypes found at high frequency in different regions and, based only on subtyping, would not appear to be geographically divergent. However, localized epidemics could result in regional diversification of subtypes, identifiable through phylogenetic analysis by the formation of closely related sequence clusters.

Remarkably, the majority of GT1b NS5B sequences (Figure 2) from the 47 patients in India (red) formed two monophyletic clusters. The GT1b NS5B phylogenetic tree also revealed clustering of sequences from other countries. Japan (n=70), Korea (n=86) and Taiwan (n=67) (in orange and yellow) formed two major clusters in which very few sequences from other countries where intermixed. Of the 39 sequences from New Zealand (light blue), 27 formed a distinct cluster. As with GT1b, the majority of GT3a NS5B sequences from patients in India (red; 18/31) formed a significant monophyletic cluster (Figure 3). Clustering of GT3a sequences from patients in Russia (gray) was also observed. Unlike India, Russian GT3a sequences form three main clusters and group with sequences from Europe and Oceania. Conversely, there is a high level of intermingling between GT1b and 3a sequences sampled from countries in North America (purple) and Europe (green). Similar clustering patterns were observed in GT1b and GT3a NS5A phylogenetic trees (data not shown), suggesting that the genetic differences observed were not specific for NS5B gene. The tight clusters of sequences could represent local HCV epidemics and diversification. However, since the sequences analyzed came from a limited number of patients recruited at clinical trial centers, more sampling will be needed in order to determine if the sequence clusters are representative of circulating virus in the aforementioned countries.

Discussion:

A comprehensive analysis of genotyping/subtyping accuracy was performed in order to understand the limitations of standard subtyping methods. The current standard is the INNO-LiPA, a line probe reverse hybridization assay that uses probes to the 5' untranslated region (UTR) and part of the core sequence to discern HCV type (12, 13). Assays that focus on a short part of the genome may not discern correct genotypes if there is high sequence homology, and are limited in identification of recombinant and/or novel strains (15, 16). Amplicon and full genome sequencing of HCV are alternate methods that provide greater resolution in the identification of genotype and subtype (16). However, it is not practical to perform sequencing in all clinical settings. Based on the geographic distribution of HCV, some regions may be impacted more by incorrect or incomplete identification by standard methods.

We found high levels of concordance in genotype assignment between INNO-LiPA and amplicon sequencing of NS3/4A, NS5A and/or NS5B, consistent with previous reports (29, 36). Overall, we observed 0.5% genotype discordance, with the highest proportion of discordant calls in GT2. Full genome sequencing analysis on GT2 discordant samples confirmed that these discordant results primarily reflect the mis-categorization of GT2-GT1 recombinant forms of HCV (16/17 cases).

6/16 characterized GT2-GT1 recombinants were RF2k/1b. RF2k/1b is known to circulate in countries in Europe, with a noted high prevalence in Georgia, and was first described in a case from St. Petersburg, Russia (15-17). The sampled viruses had indistinguishable recombination breakpoints in NS2 from the original St. Petersburg RF2k/1b. Previous phylogenetic analysis (29) has indicated that these patients were likely infected by the circulating RF2k/1b strain and are not the result of unique recombination events. Patients with RF2k/1b were sampled in Europe, including Belgium (n=2), the Netherlands (n=2) and Spain (n=1), and in the US (n=1). Belgium had its first cases of RF2k/1b reported in patients originating from Georgia and the Chechen Republic (37); however, the ethnic background of the two patients from Belgium identified in our study is unknown.

All non-RF2k/1b characterized recombinants were sampled in North America, including 6 patients with recombinant GT2b/1a and 4 patients with GT2b/1b virus. Unlike the circulating RF2k/1b, there does not appear to be a prevalent circulating form of either recombinant GT2b/1a or 2b/1b. The GT2b/1a or 2b/1b patients did not unequivocally share identical recombination breakpoints. Phylogenetic analyses (29) suggest that unique recombination events generated these viruses in some cases.

The majority of GT1 discordants were confirmed to be GT6 by full genome sequencing (6/8), indicating that the INNO-LiPA v2.0 still has difficulty in discerning between these two genotypes, though at a lower rate than the previous version (36, 38). The diversity and sequence complexity in GT6 subtypes may contribute to the continued problem in distinguishing GT1 and GT6, even with the presence of probes directed at the core.

The overall level of refinement and discordance needed in subtyping was 12.3%, however this disproportionately affected GT2, 3, 4 and 6. Strikingly, the level of novel/mixed subtypes was also highest for GT2 (3.1%), 4 (2.3%) and 6 (4.7%). The subtype diversity within each genotype appeared to be directly correlated to the higher levels of discordancy and refinement needed. The patient sequences included under novel/mixed were not fully characterized, but their occurrence in these genotypes indicates that their subtype diversity could be significantly under sampled.

The global prevalence of subtypes corroborated the observation of high diversity in GT2, 4 and 6. The high number of circulating GT2, 4 and 6 subtypes contrasts with what is seen in GT1, 3 and 5, where 1-2 subtypes predominate. Subtype diversity and prevalence varied region, as has been shown with genotype prevalence (10, 11). Interestingly, France harbored novel/mixed strains of GT1 (n=3; 0.8%), 2 (n=13; 28.3%), 4 (n=4; 3.4%), 5 (n=1; 1.8%) and 6 (n=1; 33.3%). The reason for the increased frequency of uncharacterized subtypes measured in France is unclear, though immigration patterns should be considered as a contributing factor and investigated using phylogeographic analyses. The subtype and RAS distributions determined are representative of the patients enrolled in clinical trials run at specific regional medical centers in primarily developed countries, and thus may not reflect the all of the global landscape. Migration of patients to the clinical trial centers may have impacted our reported subtype distributions as well. Greater sampling will be needed in order to determine global distribution.

The RAS analysis revealed that NS5B DAA resistance is infrequent, but NS3 and NS5A DAA RASs are detected at higher frequencies in the population. The analyzed population excluded DAA-experienced patients. In NS5B, the presence of the NS5B RAS L159F was only detected in GT1b at 8% prevalence (Table 5). L159F is associated with a lower rate of SVR in only shorter duration (<24 week) regimens containing SOF with RBV, however not in combination regimens targeting both NS5A and NS5B, with PEG-IFN, or when using recommended treatment times (39). The

NS5B RAS S282T, the only RAS with demonstrated in-vitro resistance to SOF, was not detected in any GT at baseline.

The NS5A variants at RAS position Y93 were detected in many of the subtypes sampled. The Y93H variant was found in GT1b (11%), GT3a (6%), GT4b (50%) and GT4r (13%). GT4r had a high prevalence of NS5A RASs in the 8 patients analyzed, with most patients carrying 2 NS5A RASs at baseline. In a study of lepidasvir (LDV)/SOF treatment for 12 weeks in GT4 infection, 2 of 3 patients with GT4r did not achieve SVR (40). The GT4r patient successfully treated with LDV/SOF had 2 NS5A RASs (M28I+Q30R) at baseline, similar to most GT4r patients in our analysis (50%). The two patients who failed therapy had 3 NS5A RASs (M28M/V+Q30R+L31M) at baseline. Due to the small sample size, it is unclear if failure is due to the higher number of NS5A RASs in those GT4r patients. The NS5A variants detected in GT1a and 1b this analysis corroborate results from a recent publication (41).

NS3 RASs were found in multiple subtypes in the analyzed population. D168 variants were detected in all GT3a, 3b, and 3i isolates (D168Q), 53% of GT5a (D168E) and 7% GT6a (D168E) isolates sampled, as well as in the single patients sampled representing GT1c (D168E) and GT7 (D168Q). Q80K, associated with resistance to simeprevir, was detected at a prevalence of 36% of GT1a patients, and Q80L was detected in the n=2 GT1l patients sampled. V36L, which confers resistance to telaprevir, was observed in almost all GT2-5 patients sampled. GT1c, 1e and 7 had triple combinations of NS3 RASs prevalent within the population studied. The use of combination DAA regimens, and not monotherapy, in these subtypes may lessen the impact of circulating RASs on treatment outcome.

Future work should focus on how HCV subtype prevalence changes amid ongoing cure treatments and how subtypes respond differently to current DAA regimens, especially in light of subtype-specific polymorphisms seen in our RAS analysis. Differences in regimen efficacy based on subtype would be appear to be most important for GT2, GT4 and GT6 and thus would affect countries in North and Central Africa, the Middle East and Asia more prominently. If subtypes within these genotypes are harder to treat, diagnosis and treatment optimization would be exacerbated by the lack of accurate subtyping using standard methods. Thus, continued surveillance is critical in order to develop the most effective strategies for HCV elimination worldwide.

Acknowledgements:

We extend our thanks to the patients who participated in the phase II/III clinical studies and to the clinical and research staff at Gilead.

References:

1. Shepard CW, Finelli L, Alter MJ. Global epidemiology of hepatitis C virus infection. Lancet Infect Dis. 2005;5(9):558-67.

2. Lauer GM, Walker BD. Hepatitis C virus infection. N Engl J Med. 2001;345(1):41-52.

3. Smith DB, Bukh J, Kuiken C, Muerhoff AS, Rice CM, Stapleton JT, et al. Expanded classification of hepatitis C virus into 7 genotypes and 67 subtypes: updated criteria and genotype assignment web resource. Hepatology. 2014;59(1):318-27.

4. Timm J, Roggendorf M. Sequence diversity of hepatitis C virus: implications for immune control and therapy. World J Gastroenterol. 2007;13(36):4808-17.

5. Asselah T, Boyer N, Saadoun D, Martinot-Peignoux M, Marcellin P. Direct-acting antivirals for the treatment of hepatitis C virus infection: optimizing current IFN-free treatment and future perspectives. Liver Int. 2016;36 Suppl 1:47-57.

6. Sulkowski MS, Gardiner DF, Rodriguez-Torres M, Reddy KR, Hassanein T, Jacobson I, et al. Daclatasvir plus sofosbuvir for previously treated or untreated chronic HCV infection. N Engl J Med. 2014;370(3):211-21.

7. Lawitz E, Mangia A, Wyles D, Rodriguez-Torres M, Hassanein T, Gordon SC, et al. Sofosbuvir for previously untreated chronic hepatitis C infection. N Engl J Med. 2013;368(20):1878-87.

8. Feld JJ, Jacobson IM, Hezode C, Asselah T, Ruane PJ, Gruener N, et al. Sofosbuvir and Velpatasvir for HCV Genotype 1, 2, 4, 5, and 6 Infection. N Engl J Med. 2015;373(27):2599-607.

9. Foster GR, Afdhal N, Roberts SK, Brau N, Gane EJ, Pianko S, et al. Sofosbuvir and Velpatasvir for HCV Genotype 2 and 3 Infection. N Engl J Med. 2015;373(27):2608-17.

10. Gower E, Estes C, Blach S, Razavi-Shearer K, Razavi H. Global epidemiology and genotype distribution of the hepatitis C virus infection. J Hepatol. 2014;61(1 Suppl):S45-57.

11. Messina JP, Humphreys I, Flaxman A, Brown A, Cooke GS, Pybus OG, et al. Global distribution and prevalence of hepatitis C virus genotypes. Hepatology. 2015;61(1):77-87.

12. Rehman IU, Idrees M, Ali M, Ali L, Butt S, Hussain A, et al. Hepatitis C virus genotype 3a with phylogenetically distinct origin is circulating in Pakistan. Genet Vaccines Ther. 2011;9(1):2.

13. McNaughton AL, Cameron ID, Wignall-Fleming EB, Biek R, McLauchlan J, Gunson RN, et al. Spatiotemporal Reconstruction of the Introduction of Hepatitis C Virus into Scotland and Its Subsequent Regional Transmission. J Virol. 2015;89(22):11223-32.

14. Akkarathamrongsin S, Hacharoen P, Tangkijvanich P, Theamboonlers A, Tanaka Y, Mizokami M, et al. Molecular epidemiology and genetic history of hepatitis C virus subtype 3a infection in Thailand. Intervirology. 2013;56(5):284-94.

15. Karchava M, Waldenstrom J, Parker M, Hallack R, Sharvadze L, Gatserelia L, et al. High incidence of the hepatitis C virus recombinant 2k/1b in Georgia: Recommendations for testing and treatment. Hepatol Res. 2015;45(13):1292-8.

16. Raghwani J, Thomas XV, Koekkoek SM, Schinkel J, Molenkamp R, van de Laar TJ, et al. Origin and evolution of the unique hepatitis C virus circulating recombinant form 2k/1b. J Virol. 2012;86(4):2212-20.

17. Kalinina O, Norder H, Mukomolov S, Magnius LO. A natural intergenotypic recombinant of hepatitis C virus identified in St. Petersburg. J Virol. 2002;76(8):4034-43.

18. Wyles DL. Antiviral resistance and the future landscape of hepatitis C virus infection therapy. J Infect Dis. 2013;207 Suppl 1:S33-9.

19. Paolucci S, Fiorina L, Mariani B, Gulminetti R, Novati S, Barbarini G, et al. Naturally occurring resistance mutations to inhibitors of HCV NS5A region and NS5B polymerase in DAA treatment-naive patients. Virol J. 2013;10:355.

20. Sarrazin C, Lathouwers E, Peeters M, Daems B, Buelens A, Witek J, et al. Prevalence of the hepatitis C virus NS3 polymorphism Q80K in genotype 1 patients in the European region. Antiviral Res. 2015;116:10-6.

21. Stuyver L, Rossau R, Wyseur A, Duhamel M, Vanderborght B, Van Heuverswyn H, et al. Typing of hepatitis C virus isolates and characterization of new subtypes using a line probe assay. J Gen Virol. 1993;74 (Pt 6):1093-102.

22. Stuyver L, Wyseur A, van Arnhem W, Lunel F, Laurent-Puig P, Pawlotsky JM, et al. Hepatitis C virus genotyping by means of 5'-UR/core line probe assays and molecular analysis of untypeable samples. Virus Res. 1995;38(2-3):137-57.

23. Altschul SF, Madden TL, Schaffer AA, Zhang J, Zhang Z, Miller W, et al. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic Acids Res. 1997;25(17):3389-402.

24. Hedskog C, Chodavarapu K, Ku KS, Xu S, Martin R, Miller MD, et al. Genotype- and Subtype-Independent Full-Genome Sequencing Assay for Hepatitis C Virus. J Clin Microbiol. 2015;53(7):2049-59.

25. Thompson JD, Higgins DG, Gibson TJ. CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. Nucleic acids research. 1994;22(22):4673-80.

26. DJ Z. Genetic algorithm approaches for the phylogenetic analysis of large biological sequence datasets under the maximum likelihood criterion. Ph.D. dissertation: The University of Texas at Austin; 2006.

27. Anisimova M, Gascuel O. Approximate likelihood-ratio test for branches: A fast, accurate, and powerful alternative. Syst Biol. 2006;55(4):539-52.

28. Guindon S, Dufayard JF, Lefort V, Anisimova M, Hordijk W, Gascuel O. New algorithms and methods to estimate maximum-likelihood phylogenies: assessing the performance of PhyML 3.0. Syst Biol. 2010;59(3):307-21.

29. Hedskog C, Doehle B, Chodavarapu K, Gontcharova V, Crespo Garcia J, De Knegt R, et al. Characterization of hepatitis C virus intergenotypic recombinant strains and associated virological response to sofosbuvir/ribavirin. Hepatology. 2015;61(2):471-80.

30. Schreiber J, McNally J, Chodavarapu K, Svarovskaia E, Moreno C. Treatment of a patient with genotype 7 HCV infection with sofosbuvir and velpatasvir. Hepatology. 2016.

31. Frank C, Mohamed MK, Strickland GT, Lavanchy D, Arthur RR, Magder LS, et al. The role of parenteral antischistosomal therapy in the spread of hepatitis C virus in Egypt. Lancet. 2000;355(9207):887-91.

32. Lin MV, Chung R. Recent FDA approval of sofosbuvir and simeprevir. Implications for current HCV treatment. Clinical Liver Disease. 2014;3(3):65-8.

33. Kumada H, Suzuki Y, Ikeda K, Toyota J, Karino Y, Chayama K, et al. Daclatasvir plus asunaprevir for chronic HCV genotype 1b infection. Hepatology. 2014;59(6):2083-91.

34. Nakamoto S, Kanda T, Wu S, Shirasawa H, Yokosuka O. Hepatitis C virus NS5A inhibitors and drug resistance mutations. World J Gastroenterol. 2014;20(11):2902-12.

35. Lam AM, Espiritu C, Bansal S, Micolochick Steuer HM, Niu C, Zennou V, et al. Genotype and subtype profiling of PSI-7977 as a nucleotide inhibitor of hepatitis C virus. Antimicrob Agents Chemother. 2012;56(6):3359-68.

36. Bouchardeau F, Cantaloube JF, Chevaliez S, Portal C, Razer A, Lefrere JJ, et al. Improvement of hepatitis C virus (HCV) genotype determination with the new version of the INNO-LiPA HCV assay. J Clin Microbiol. 2007;45(4):1140-5.

37. De Keukeleire S, Descheemaeker P, Reynders M. Diagnosis of hepatitis C virus genotype 2k/1b needs NS5B sequencing. Int J Infect Dis. 2015;41:1-2.

38. Noppornpanth S, Sablon E, De Nys K, Truong XL, Brouwer J, Van Brussel M, et al. Genotyping hepatitis C viruses from Southeast Asia by a novel line probe assay that simultaneously detects core and 5' untranslated regions. J Clin Microbiol. 2006;44(11):3969-74.

39. Svarovskaia ES, Gane E, Dvory-Sobol H, Martin R, Doehle B, Hedskog C, et al. L159F and V321A Sofosbuvir-Associated Hepatitis C Virus NS5B Substitutions. J Infect Dis. 2016;213(8):1240-7.

40. Abergel A, Metivier S, Samuel D, Jiang D, Kersey K, Pang PS, et al. Ledipasvir Plus Sofosbuvir for 12 Weeks in Patients With Hepatitis C Genotype 4 Infection. Hepatology. 2016.

41. Zeuzem S, Mizokami M, Pianko S, Mangia A, Han KH, Martin R, et al. NS5A Resistance-Associated Substitutions in Patients with Genotype 1 Hepatitis C Virus: Prevalence and Effect on Treatment Outcome. J Hepatol. 2017.

Figure Legends:

Figure 1. Subtype distributions of HCV genotypes 1-6 determined using sequencing-based analysis.

The proportions of subtypes circulating for GT1 (A), GT2 (B), GT3 (C), GT4 (D), GT5 (E), and GT6 (F) are plotted for each country sampled. The color key for different subtypes is shown to the right of each figure. The novel/mixed category refers to sequences assayed that did not sufficiently match a characterized subtype. Countries are organized by geographic region, moving left to right: North America, Europe, Asia, Oceania, and/or Africa.

Figure 2. Maximum likelihood phylogenetic tree of genotype 1b NS5B sequences.

621 GT1b sequences were included in the maximum-likelihood tree and branches with statistical significance ($p<0.001$) are noted with an asterisk. One significant cluster of sequences from India (red) contained 21/47 sequences and the other contained 11/47 sequences. The two significant branches with sequences from Japan, Korea and Taiwan (yellow and orange) contained 223 sequences. 27 sequences from New Zealand (blue) formed a significant cluster.

Figure 3. Maximum likelihood phylogenetic tree of genotype 3a NS5B sequences.

351 sequences GT3a sequences were included in the maximum-likelihood tree and branches with statistical significance ($p<0.001$) are noted with an asterisk. 18/31 sequences from India (red) formed a cluster and the 45 sequences from Russia (gray) formed three main clusters (Cluster 1 Russian sequences n=16, England n=1; Cluster 2

Russian sequences n=9; Cluster 3 Russian sequences n=19, Estonia n=3, Australia n=1).

Table 1. Summary of Concordant and Discordant Results by HCV Genotype and Subtype

HCV Genotype (by INNO-LiPA, 5'UTR and Core detection)	Patients	HCV Genotype (by sequencing NS3/4A, NS5A and/or NS5B)				HCV Subtype (by sequencing NS3/4A, NS5A and/or NS5B)							
		Concordant		Discordant		Concordant		Refined		Discordant		Novel/mixed	
	n	n	%	n	%	n	%	n	%	n	%	n	%
1	5701	5691	99.8%	10[a]	0.2%	5576	97.8%	80	1.4%	35	0.6%	10	0.2%
2	950	921	96.9%	29[b]	3.1%	485	51.1%	394	41.5%	42	4.4%	29	3.1%
3	1841	1840	99.9%	1[c]	0.1%	1659	90.1%	174	9.5%	6	0.3%	2	0.1%
4	309	308	99.7%	1	0.3%	18	5.8%	251	81.2%	33	10.7%	7	2.3%
5	37	37	100.0%	0	0.0%	37	100.0%	NA	NA	NA	NA	NA	NA
6	107	107	100.0%	0	0.0%	10	9.3%	83	77.6%	9	8.4%	5	4.7%
Total	**8945**	**8904**	**99.5%**	**41[d]**	**0.5%**	**7785**	**87.0%**	**982**	**11.0%**	**125**	**1.4%**	**53**	**0.6%**

Concordant: Genotype/subtype result from INNO-LiPA was an exact match to the genotype/subtype determined by sequencing.
Discordant: Genotype/subtype result from INNO-LiPA did not match the genotype/subtype determined by sequencing.
Refined: The category of genotype/subtype from INNO-LiPA (e.g. 4a/c/d or 6c-l) was correct but not as specific as the genotype/subtype determined by sequencing.
Novel/mixed: The genotype/subtype could not be specifically determined by sequencing and was interpreted to indicate a novel or mixed subtype.

[a] 6/6 INNO-LiPA genotype 1 discordant samples tested by full genome sequencing were confirmed to be genotype 6.
[b] 16/17 INNO-LiPA genotype 2 discordant samples tested by full genome sequencing were confirmed to be genotype 2/1 recombinants and 1/17 was confirmed to be genotype 7.
[c] The INNO-LiPA genotype 3 discordant sample tested by full genome sequencing was confirmed to be genotype 1.
[d] 4/41 samples had inconclusive INNO-LiPA results and were additionally genotyped using TRUGENE HCV 5'NC kit (Siemens Healthcare Diagnostics).

Table 2. Prevalence of HCV Subtypes by Geographic Region

Total (n=12,615)	North America (n=7,891)		Europe (n=2,166)		Asia (n=977)		Oceania (n=1,474)		Africa (n=107)	
Subtype	**n**	**%**	**n**	**%**	**n**	**%**	**n**	**%**	**n**	**%**
1a	4265	76.4	552	49.9	26	4.1	451	78.0		
1b	1295	23.2	536	48.4	612	95.6	126	21.8		
1c	2	0.0	0	0.0	0	0.0	0	0.0		
1e	0	0.0	5	0.5	0	0.0	0	0.0		
1g	1	0.0	0	0.0	0	0.0	1	0.2		
1h	1	0.0	1	0.1	0	0.0	0	0.0		
1l	2	0.0	3	0.3	0	0.0	0	0.0		
Genotype 1 Novel/mixed	4	0.1	5	0.5	2	0.3	0	0.0		
Recombinant 2/1*	11	0.2	5	0.5	0	0.0	0	0.0		
2a	148	15.5	26	16.3	35	35.4	35	33.3		
2b	787	82.7	43	26.9	59	59.6	64	61.0		
2c	5	0.5	54	33.8	5	5.1	6	5.7		
2d	0	0.0	1	0.6	0	0.0	0	0.0		
2i	2	0.2	10	6.3	0	0.0	0	0.0		
2j	1	0.1	1	0.6	0	0.0	0	0.0		
2k	1	0.1	3	1.9	0	0.0	0	0.0		
Genotype 2 Novel/mixed	8	0.8	22	13.8	0	0.0	0	0.0		
3a	1059	98.7	623	98.9	79	75.2	752	98.7		
3b	7	0.7	3	0.5	12	11.4	6	0.8		
3g	3	0.3	1	0.2	4	3.8	0	0.0		
3h	0	0.0	2	0.3	0.0	0.0	0	0.0		
3i	4	0.4	0	0.0	8.0	7.6	0	0.0		
3k	0	0.0	1	0.2	0	0.0	4	0.5		
Genotype 3 Novel/mixed	0	0.0	0	0.0	2	1.9	0	0.0		
4a	182	82.4	81	42.2			5	45.5	86	84.3
4b	0	0.0	10	5.2			0	0.0	0	0.0
4c	0	0.0	3	1.6			0	0.0	0	0.0
4d	8	3.6	65	33.9			5	45.5	0	0.0
4f	0	0.0	4	2.1			0	0.0	0	0.0
4k	2	0.9	1	0.5			0	0.0	0	0.0
4l	3	1.4	0	0.0			0	0.0	1	1.0
4m	1	0.5	1	0.5			0	0.0	0	0.0
4n	7	3.2	3	1.6			1	9.1	0	0.0
4o	13	5.9	6	3.1			0	0.0	5	4.9
4r	1	0.5	8	4.2			0	0.0	0	0.0
4t	0	0.0	1	0.5			0	0.0	0	0.0

Genotype 4 Novel/mixed	4	1.8	9	4.7			0	0.0	10	9.8
5a	8	100.0	69	98.6					5	100.0
Genotype 5 Novel/mixed	0	0.0	1	1.4					0	0.0
6a	21	37.5	2	33.3	26	74.3	8	33.3		
6e	18	32.1	0	0.0	0	0.0	5	20.8		
6f	0	0.0	0	0.0	1	2.9	0	0.0		
6g	0	0.0	0	0.0	5	14.3	0	0.0		
6h	3	5.4	0	0.0	0	0.0	0	0.0		
6i	0	0.0	0	0.0	1	2.9	0	0.0		
6l	5	8.9	1	16.7	0	0.0	2	8.3		
6m	1	1.8	0	0.0	0	0.0	2	8.3		
6n	1	1.8	1	16.7	2	5.7	0	0.0		
6o	1	1.8	0	0.0	0	0.0	0	0.0		
6p	0	0.0	0	0.0	0	0.0	2	8.3		
6q	2	3.6	1	16.7	0	0.0	2	8.3		
6r	0	0.0	0	0.0	0	0.0	1	4.2		
Genotype 6 Novel/mixed	4	7.1	1	16.7	0	0.0	2	8.3		
GT7			1	100.0						

North America: US, Canada and Puerto Rico
Europe: Great Britain, Ireland, Sweden, Germany, Netherlands, Belgium, France, Spain, Italy, Switzerland, Poland and Estonia
Asia: Russia, India, China, Hong Kong, Taiwan, Korea, Japan and Thailand
Oceania: Australia and New Zealand
Africa: Egypt and South Africa
Novel/mixed: The genotype/subtype could not be specifically determined by sequencing and was interpreted to indicate a novel or mixed subtype.

*Excluded 13 putative GT2/1 recombinants not confirmed by full genome sequence.

Table 3. NS3 RASs in HCV GT1-7 Subtypes

		NS3 RAS POSITION[a,b,c]								
Subtype	**Patients (n)***	**V36[d]**	**Q41**	**T54**	**V55**	**Q80**	**S122**	**R155**	**A156**	**D168**
1a	1474	-[e]	-	-	-	**Q (64%)** K (36%)	**S (94%)** G (8%)	**R (98%)**	-	-
1b	583	-	-	-	-	-	**S (87%)** T (7%), G (6%), N (5%)	-	-	-
1c	1	NA	NA	NA	NA	NA	NA	**D (100%)**	**T (100%)**	**E (100%)**
1e	2	**L (100%)**	-	**S (100%)**	-	-	**N (100%)**	-	-	-
1g	0									
1h	0									
1l	2	-	-	-	-	**L (100%)**	-	-	-	-
2k-1b	4	-	-	-	-	-	**S (50%)** G (25%), N (25%)	-	-	-
2a	31	**L (94%)** M (6%)	-	-	-	**G (100%)**	**K (97%)**	-	-	-
2b	59	**L (100%)**	-	-	-	**G (100%)**	**R (100%)**	-	-	-
2c	10	**L (100%)**	-	-	-	**G (100%)**	**R (100%)**	-	-	-
2d	0									
2e	0									
2i	2	**L (100%)**	-	-	-	**G (100%)**	**R (100%)**	-	-	-
2j	1	**L (100%)**	-	-	-	**G (100%)**	**R (100%)**	-	-	-
2k	0									
2r	0									
3a	383	**L (100%)**	-	-	-	-	-	-	-	**Q (100%)**
3b	2	**L (100%)**	-	-	-	-	-	-	-	**Q (100%)**
3g	0									
3h	0									
3i	1	**L (100%)**	-	-	-	-	-	-	-	**Q (100%)**
3k	0									

4a	50	**L (100%)**	-	-	-	-	**T (100%)**	-	-	-
4b	0									
4c	1	**L (100%)**	-	-	-	-	**T (100%)**	-	-	-
4d	12	**L (100%)**	-	-	-	-	**T (92%)** S (8%)	-	-	-
4f	0									
4k	1	**L (100%)**	-	-	-	-	**T (100%)**	-	-	-
4l	0									
4m	0									
4n	2	**L (100%)**	-	-	-	-	**T (100%)**	-	-	-
4o	1	**L (100%)**	-	-	-	-	**T (100%)**	-	-	-
4r	0									
4t	0									
5a	15	**L (100%)**	-	-	-	**K (100%)**	**A (53%)** T (47%), G (7%), V (7%)	-	-	**D (60%)** **E (53%)**
6a	15	-	-	-	-	**K (100%)**	**N (100%)**	-	-	**D (93%)** E (7%)
6d	0									
6e	11	-	-	-	-	-	**T (82%)** A (18%), S (9%)	-	-	-
6f	0									
6g	0									
6h	0									
6k	0									
6l	3	-	-	-	-	-	**S (100%)** N (33%)	-	-	-
6m	1	-	-	-	-	-	**T (100%)**	-	-	-
6n	1	-	-	-	-	-	-	-	-	-
6o	0									
6p	0									
6q	0									
6r	0									
7	1	**L (100%)**	-	-	**A (100%)**	**D (100%)**	**K (100%)**	-	-	**Q (100%)**

[a] RASs reported with ≥5% prevalence within each analyzed subtype population.
[b] RASs detected with ≥15% abundance in patient virus sequence.
[c] RASs carried by the majority of the subtype population (≥50% prevalence) are in bold. RASs held by a minority (<50%) of the subtype population are not in bold and are listed in order of prevalence.
[d] Reference amino acid at RAS site based on GT1a H77 reference sequence is not associated with resistance. Any changes from this reference are shown.
[e] Dash (-) indicates no change from reference and no minority substitutions detected in analysis.
* Patient n represents all patients analyzed across all positions. Actual analyzed number may be lower depending upon position.
NA = not available; a partial sequence did not overlap the analyzed RAS position.

Table 4. NS5A RASs in HCV GT1-7 Subtypes

		NS5A RAS POSITION[a,b,c]								
Subtype	**Patients (n)***	**K24[d]**	**M28**	**Q30**	**L31**	**P32**	**S38**	**H58**	**A92**	**Y93**
1a	3532	-[e]	**M (96%)** V (6%)	-	-	-	-	**H (94%)** P (6%)	-	-
1b	1972	**Q (97%)**	**L (98%)**	**R (92%)** Q (9%)	**L (96%)**	-	-	**P (93%)** S (5%)	-	**Y (96%)** H (11%)
1c	2	-	**V (100%)**	-	-	-	-	**P (100%)**	-	-
1e	5	**Q (60%)** K (20%), R (20%)	-	**Q (60%)** L (20%), R (20%)	**L (60%)** M (40%)	-	-	**P (100%)**	**T (60%)** A (20%), V (20%)	-
1g	2	**R (50%)** **S (50%)**	**L (100%)**	-	-	-	-	**P (100%)**	-	**F (100%)**
1h	2	**Q (100%)**	**L (100%)**	**R (100%)**	-	-	-	**P (100%)**	-	-
1l	3	**G (67%)** **S (67%)**	-	**R (67%)** Q (33%)	**M (100%)**	-	-	**P (100%)**	-	-
2k-1b	4	**Q (100%)**	**L (100%)**	**R (100%)**	-	-	-	**P (100%)**	**V (75%)** A (25%)	-
2a	106	**T (80%)** A (19%)	**F (95%)**	**K (97%)**	**M (97%)**	-	-	**P (95%)** S (6%)	**C (95%)** S (5%)	-
2b	441	**S (100%)**	**L (98%)**	**K (100%)**	**L (67%)** M (39%)	-	-	**P (96%)** S (5%)	**C (100%)**	-
2c	27	**S (100%)**	**F (59%)** C (41%)	**K (96%)** R (11%)	**L (89%)** M (11%)	-	-	**P (93%)**	**C (100%)**	-
2d	2	**S (100%)**	**L (100%)**	**K (100%)**	**L (50%)** **M (50%)**	-	-	**P (100%)**	**C (100%)**	-
2e	2	**S (100%)**	**L (50%)** **F (50%)**	**K (100%)**	**L (50%)** **M (50%)**	-	-	**P (100%)**	**C (100%)**	-
2i	11	**S (100%)**	**F (100%)**	**K (100%)** R (9%)	**M (100%)**	-	-	**P (100%)**	**C (100%)**	-
2j	2	**S (100%)**	**F (100%)**	**K (50%)** **R (50%)**	**M (100%)**	-	-	**P (100%)**	**C (100%)**	-
2k	9	**S (100%)** T (11%)	**L (67%)** F (44%)	**K (89%)** S (11%)	**M (67%)** L (22%), I (11%)	-	-	**P (89%)** S (11%)	**C (100%)** S (11%)	-
2r	2	**S (100%)**	**L (50%)** **C (50%)**	**K (100%)**	-	-	-	**P (100%)**	**C (100%)**	-
3a	1263	**S (99%)**	-	**A (91%)** K (5%)	-	-	-	**P (97%)**	**E (100%)**	**Y (97%)** H (6%)
3b	25	**S (92%)** A (8%)	-	**K (96%)**	**M (92%)** L (8%)	-	-	**P (100%)**	**E (100%)**	-
3g	5	**A (60%)** **S (40%)**	-	**K (100%)**	**M (80%)** V (20%)	-	-	**P (100%)**	**E (100%)**	-

3h	1	**S (100%)**	-	**A (100%)**	-	-	-	**P (100%)**	**E (100%)**	-
3i	12	**S (100%)**	-	**K (100%)** R (8%)	-	-	-	**P (100%)**	**E (100%)**	-
3k	3	**G (100%)**	**M (67%)** L (33%)	**K (100%)**	**M (100%)**	-	-	**P (100%)**	**E (100%)**	-
4a	166	-	**L (82%)** M (18%)	**L (87%)** R (12%)	**M (100%)**	-	-	**P (96%)**	-	-
4b	4	-	**L (50%)** **M (50%)**	**S (50%)** I (25%), L (25%)	**M (100%)**	-	-	**P (50%)** S (25%), T (25%)	-	**H (50%)** **Y (50%)**
4c	3	-	**L (100%)**	**R (100%)**	**M (67%)** L (33%)	-	-	**P(67%)** T (33%)	-	-
4d	66	-	**L (98%)**	**R (98%)**	**M (97%)** L (5%)	-	-	**P (62%)** T (30%), A (6%)	-	-
4f	4	-	**L (100%)**	**R (100%)**	**M (100%)**	-	-	**P (100%)**	-	-
4k	3	-	**L (100%)**	**R (100%)**	**L (67%)** M (33%)	-	-	**P (100%)**	-	-
4l	1	-	**L (100%)**	**R (100%)**	**M (100%)**	-	-	**P (100%)**	-	-
4m	1	-	**L (100%)**	**S (100%)**	**M (100%)**	-	-	**P (100%)**	-	-
4n	9	-	**L (100%)**	**R (100%)**	**M (100%)**	-	-	**T (78%)** P (22%), N (11%)	-	-
4o	12	-	-	**T (92%)** S (8%)	**M (92%)** V (8%)	-	-	**P (100%)**	-	-
4r	8	-	M (38%), V (38%), I (13%), L (13%)	**R (100%)**	**L (75%)** M (25%)	-	-	**P (100%)**	-	**Y (100%)** H (13%)
4t	1	-	**L (100%)**	**R (100%)**	**M (100%)**	-	-	**P (100%)**	-	-
5a	77	**Q (100%)**	**L (100%)**	-	-	-	-	**P (97%)**	-	-
6a	55	**Q (89%)** K (9%), R (7%)	**L (65%)** F (45%)	**R (100%)**	-	-	-	**T (96%)**	-	**T (98%)**
6d	2	-	**V (100%)**	**A (100%)**	-	-	-	**P (100%)**	-	**T (100%)**
6e	24	**K (83%)** R (17%)	**V (75%)** M (33%)	**S (96%)**	-	-	-	**P (96%)**	-	**T (96%)** S (8%)
6f	1	-	**M (100%)** **V (100%)**	**A (100%)**	-	-	-	**P (100%)**	-	**T (100%)** **S (100%)**
6g	3	**R (100%)**	**L (100%)**	**N (67%)** **S (67%)**	-	-	-	**P (67%)** S (33%)	-	**T (100%)**
6h	2	-	**V (100%)**	**A (100%)**	-	-	-	**P (100%)**	-	**T (100%)**
6k	1	-	**V (100%)**	**A (100%)**	-	-	-	**P (100%)**	-	**T (100%)**
6l	8	-	**V (100%)**	**A (88%)** T (13%)	-	-	-	**P (100%)**	-	**T (100%)**
6m	3	-	**V (100%)**	**S (100%)**	-	-	-	**T (100%)**	-	**S (100%)**
6n	2	-	**M (50%)** **V (50%)**	**S (100%)**	-	-	-	**T (100%)**	-	**S (100%)**

6o	1	-	**L (100%)**	**A (100%)**	-	-	-	**A (100%)**	-	**T (100%)**
6p	2	-	**V (100%)**	**S (100%)**	-	-	-	**P (100%)**	**A (100%)** P (50%)	**T (100%)**
6q	4	-	**V (75%)** **M (50%)**	**S (100%)**	-	-	-	**P (75%)** L (25%)	-	**T (100%)**
6r	1	-	**A (100%)**	**A (100%)**	-	-	-	**P (100%)**	-	**T (100%)**
7	1	-	**L (100%)**	**L (100%)**	-	-	-	**P (100%)**	-	**H (100%)**

[a] RASs reported with ≥5% prevalence within each analyzed subtype population.
[b] RASs detected with ≥15% abundance in patient virus sequence.
[c] RASs carried by the majority of the subtype population (≥50% prevalence) are in bold. RASs held by a minority (<50%) of the subtype population are not in bold and are listed in order of prevalence.
[d] Reference amino acid at RAS site based on GT1a H77 reference sequence is not associated with resistance. Any changes from this reference are shown.
[e] Dash (-) indicates no change from reference and no minority substitutions detected in analysis.
* Patient n represents all patients analyzed across all positions. Actual analyzed number may be lower depending upon position.

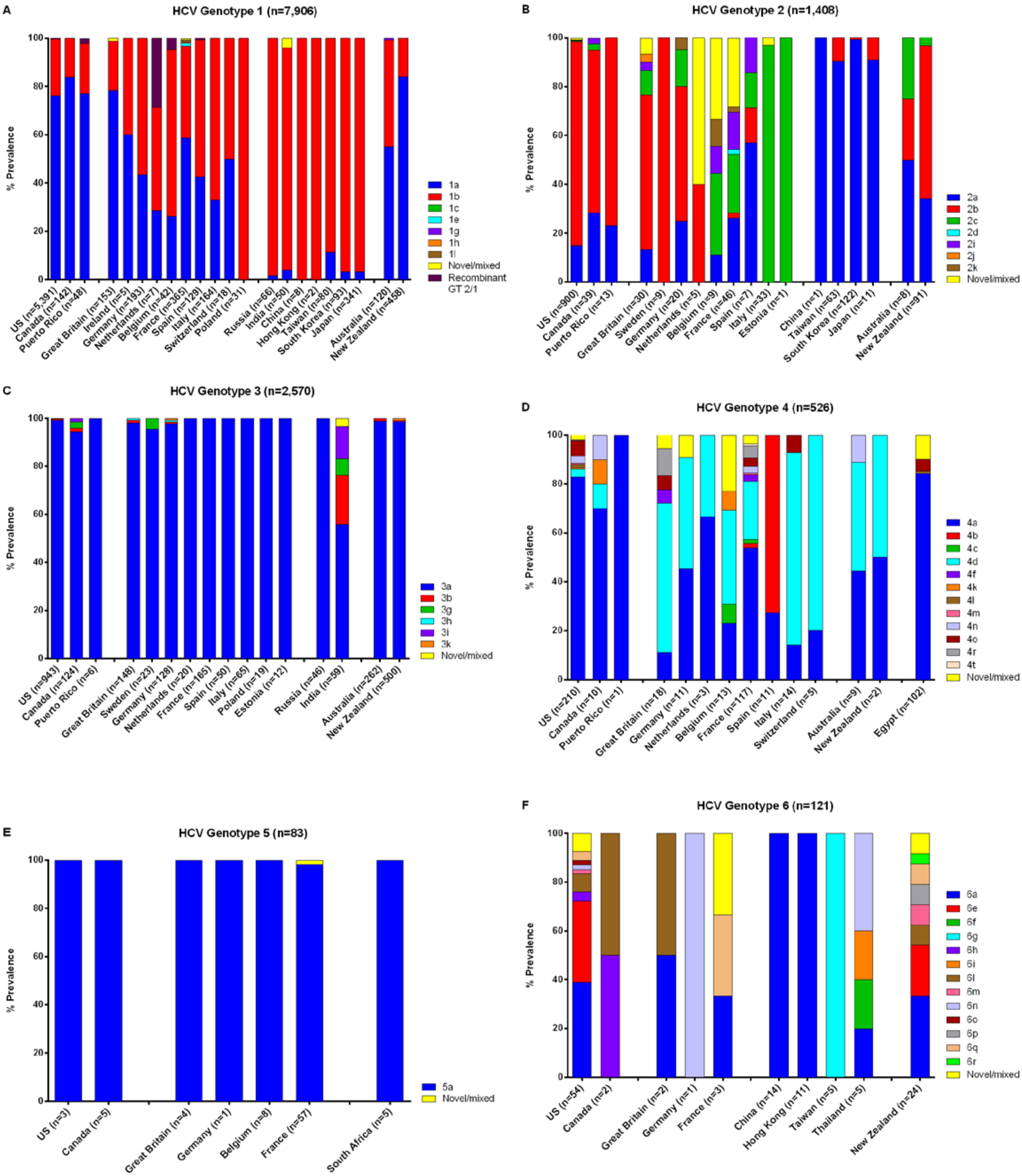
Table 5. NS5B RASs in HCV GT1-7 Subtypes

Subtype	Patients (n)*	NS5B RAS POSITION[a,b,c] S96[d]	N142	L159	E237	S282	C289	L320	V321
1a	4268	-[e]	-	-	-	-	-	-	-
1b	2243	-	**N (96%)** S (5%)	**L (92%)** F (8%)	-	-	-	-	-
1c	3	-	-	-	-	-	-	-	-
1e	6	-	-	-	-	-	-	-	-
1g	2	-	-	-	-	-	-	-	-
1h	1	-	-	-	-	-	-	-	-
1l	3	-	-	-	-	-	-	-	-
2k-1b	6	-	-	-	-	-	-	-	-
2a	399	-	-	-	-	-	**M (96%)**	-	-
2b	869	-	-	-	-	-	**M (92%)** I (9%)	-	-
2c	65	-	-	-	-	-	**M (97%)**	-	-
2d	3	-	-	-	-	-	**M (100%)**	-	-
2e	2	-	-	-	-	-	**M (100%)**	-	-
2i	12	-	-	-	-	-	**M (100%)**	-	-
2j	2	-	-	-	-	-	**M (100%)**	-	-
2k	11	-	-	-	-	-	**M (91%)** L (9%)	-	**V (91%)** I (9%)
2r	2	-	-	-	-	-	**M (100%)**	-	-
3a	2434	-	-	-	-	-	**F (100%)**	-	-
3b	27	-	**N (74%)** S (22%)	-	-	-	**F (100%)**	-	-
3g	7	-	-	-	-	-	**F (100%)**	-	-
3h	2	NA	NA	NA	-	-	**L (100%)**	-	-
3i	12	-	-	-	-	-	**F (100%)**	-	-
3k	4	NA	NA	NA	**H (50%)** E (25%), L (25%)	-	**F (100%)**	-	-

4a	349	-	-	-	**E (81%)** A (12%), G (6%)	-	**F (99%)**	-	-
4b	3	-	-	-	-	-	**F (100%)**	-	-
4c	5	-	-	-	**E (80%)** A (20%)	-	**F (100%)**	-	-
4d	86	-	-	-	-	-	**F (100%)**	-	-
4f	4	-	-	-	-	-	**F (100%)**	-	-
4k	2	-	-	-	-	-	**F (100%)**	-	-
4l	14	-	-	-	**E (93%)** G (7%)	-	**F (100%)**	-	-
4m	2	NA	NA	NA	-	-	**F (100%)**	-	-
4n	11	-	-	-	-	-	**F (100%)**	-	-
4o	24	-	-	-	-	-	**F (100%)**	-	-
4r	8	-	**N (75%)** S (25%)	-	-	-	**F (100%)**	-	**I (75%)** V (25%)
4t	1	-	-	-	-	-	**F (100%)**	-	-
5a	76	-	-	-	-	-	**M (96%)**	-	-
6a	55	-	-	-	**N (100%)**	-	**M (96%)**	-	-
6d	2	NA	NA	NA	**H (100%)**	-	**L (100%)**	-	-
6e	23	-	-	-	**H (52%)** R (43%)	-	**L (100%)**	-	-
6f	1	-	-	-	**H (100%)**	-	**L (100%)**	-	-
6g	5	NA	NA	NA	-	-	**M (100%)**	-	-
6h	2	-	-	-	**Q (100%)**	-	**M (100%)**	-	-
6k	1	-	-	-	**R (100%)**	-	**L (100%)**	-	-
6l	8	-	-	-	**R (100%)**	-	**L (100%)**	-	-
6m	3	-	-	-	**H (100%)**	-	**M (100%)**	-	-
6n	2	-	-	-	**H (100%)**	-	**L (100%)**	-	-
6o	1	-	-	-	**H (100%)**	-	**L (100%)**	-	-
6p	2	-	-	-	**H (100%)**	-	**I (50%)** **L (50%), M (50%)**	-	-
6q	5	-	-	-	**H (100%)**	-	**L (100%)**	-	-
6r	1	NA	NA	NA	-	-	**L (100%)**	-	-
7	1	-	-	-	**C (100%)**	-	-	-	-

[a] RASs reported with ≥5% prevalence within each analyzed subtype population.
[b] RASs detected with ≥15% abundance in patient virus sequence.
[c] RASs carried by the majority of the subtype population (≥50% prevalence) are in bold. RASs held by a minority (<50%) of the subtype population are not in bold and are listed in order of prevalence.
[d] Reference amino acid at RAS site based on GT1a H77 reference sequence is not associated with resistance. Any changes from this reference are shown.
[e] Dash (-) indicates no change from reference and no minority substitutions detected in analysis.
* Patient n represents all patients analyzed across all positions. Actual analyzed number may be lower depending upon position.
NA = not available; a partial sequence did not overlap the analyzed RAS position.

Figure 1.



A
HCV Genotype 1 (n=7,906)
% Prevalence
0
20
40
60
80
100
1a
1b
1c
1e
1g
1h
1l
Novel/mixed
Recombinant GT 2/1
US (n=5,391)
Canada (n=142)
Puerto Rico (n=48)
Great Britain (n=153)
Ireland (n=5)
Germany (n=193)
Netherlands (n=7)
Belgium (n=42)
France (n=365)
Spain (n=129)
Italy (n=164)
Switzerland (n=18)
Poland (n=31)
Russia (n=66)
India (n=50)
China (n=8)
Hong Kong (n=2)
Taiwan (n=80)
South Korea (n=93)
Japan (n=341)
Australia (n=120)
New Zealand (n=458)
B
HCV Genotype 2 (n=1,408)
% Prevalence
2a
2b
2c
2d
2i
2j
2k
Novel/mixed
US (n=900)
Canada (n=39)
Puerto Rico (n=13)
Great Britain (n=30)
Sweden (n=9)
Germany (n=20)
Netherlands (n=5)
Belgium (n=9)
France (n=46)
Spain (n=7)
Italy (n=33)
Estonia (n=1)
China (n=1)
Taiwan (n=63)
South Korea (n=122)
Japan (n=11)
Australia (n=8)
New Zealand (n=91)
C
HCV Genotype 3 (n=2,570)
% Prevalence
3a
3b
3g
3h
3i
3k
Novel/mixed
US (n=943)
Canada (n=124)
Puerto Rico (n=6)
Great Britain (n=148)
Sweden (n=23)
Germany (n=128)
Netherlands (n=20)
France (n=165)
Spain (n=50)
Italy (n=65)
Poland (n=19)
Estonia (n=12)
Russia (n=46)
India (n=59)
Australia (n=262)
New Zealand (n=500)
D
HCV Genotype 4 (n=526)
% Prevalence
4a
4b
4c
4d
4f
4k
4l
4m
4n
4o
4r
4t
Novel/mixed
US (n=210)
Canada (n=10)
Puerto Rico (n=1)
Great Britain (n=18)
Germany (n=11)
Netherlands (n=3)
Belgium (n=13)
France (n=117)
Spain (n=11)
Italy (n=14)
Switzerland (n=5)
Australia (n=9)
New Zealand (n=2)
Egypt (n=102)
E
HCV Genotype 5 (n=83)
% Prevalence
5a
Novel/mixed
US (n=3)
Canada (n=5)
Great Britain (n=4)
Germany (n=1)
Belgium (n=8)
France (n=57)
South Africa (n=5)
F
HCV Genotype 6 (n=121)
% Prevalence
6a
6e
6f
6g
6h
6i
6l
6m
6n
6o
6p
6q
6r
Novel/mixed
US (n=54)
Canada (n=2)
Great Britain (n=2)
Germany (n=1)
France (n=3)
China (n=14)
Hong Kong (n=11)
Taiwan (n=5)
Thailand (n=5)
New Zealand (n=24)

Figure 2.

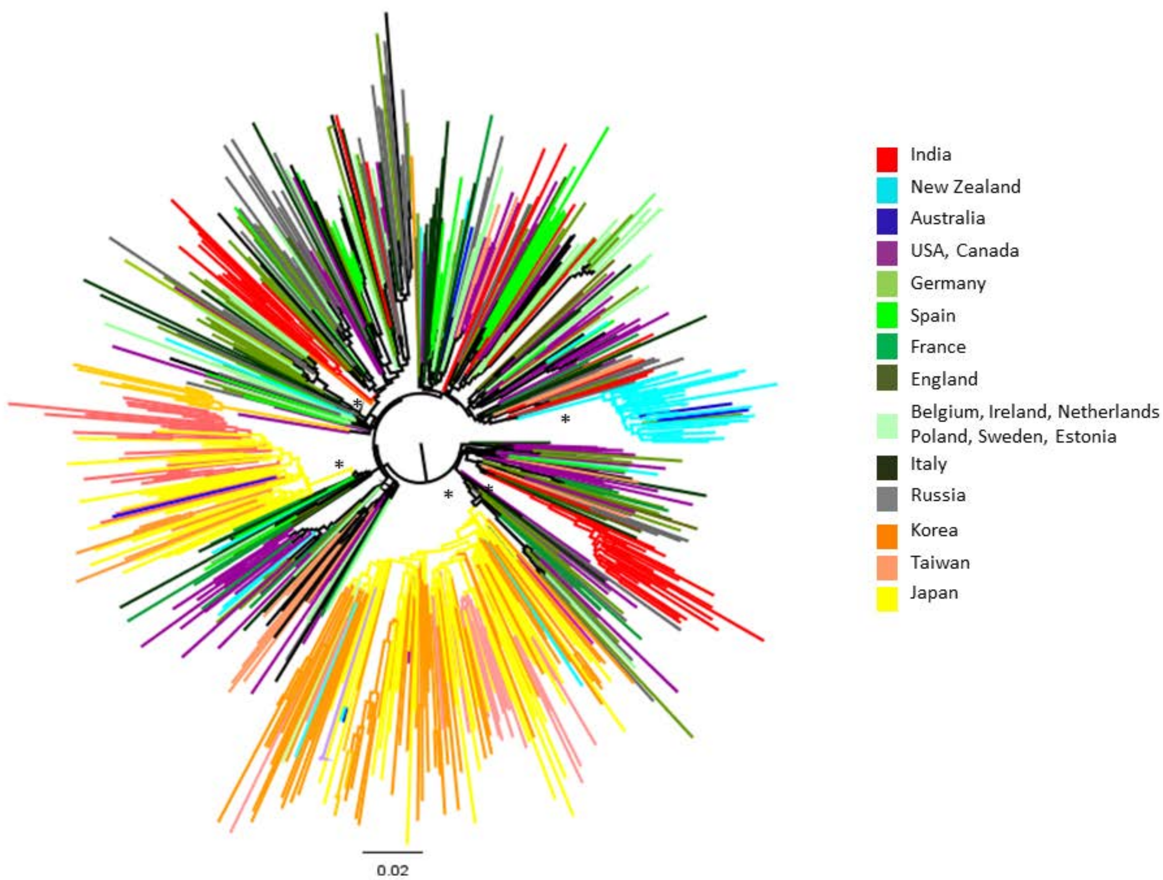
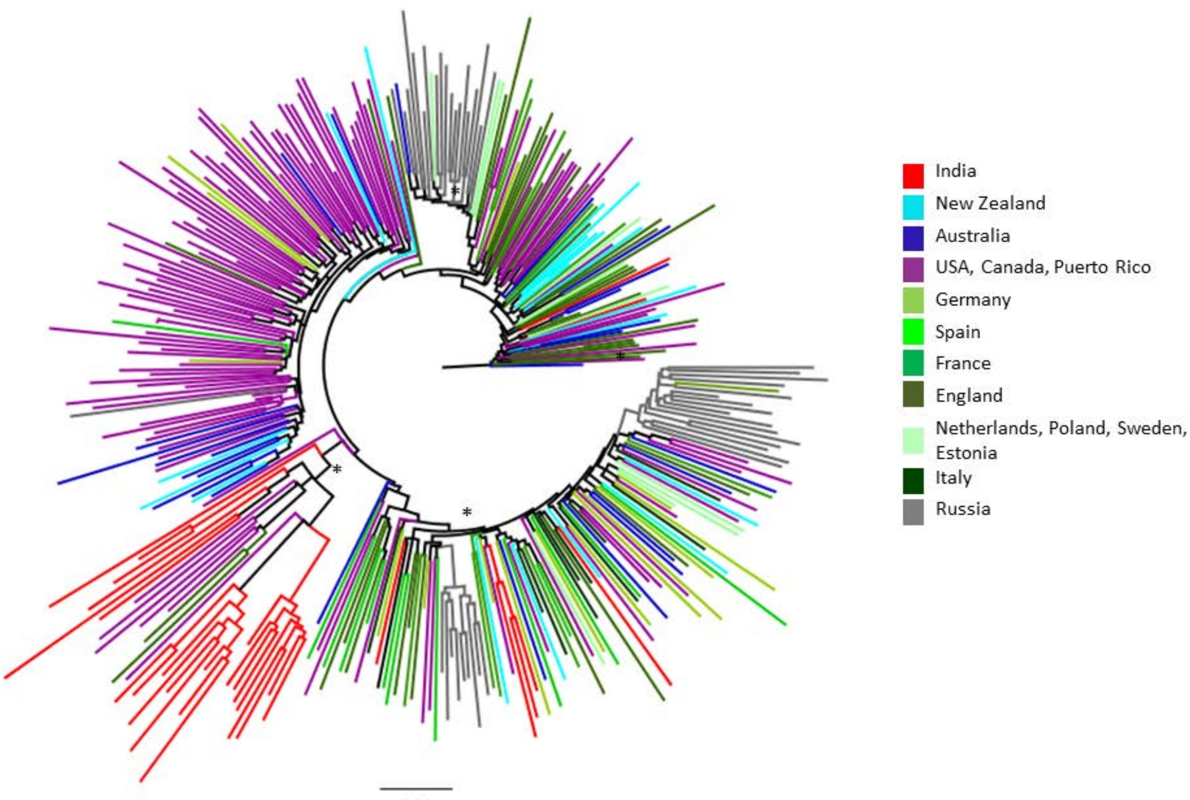
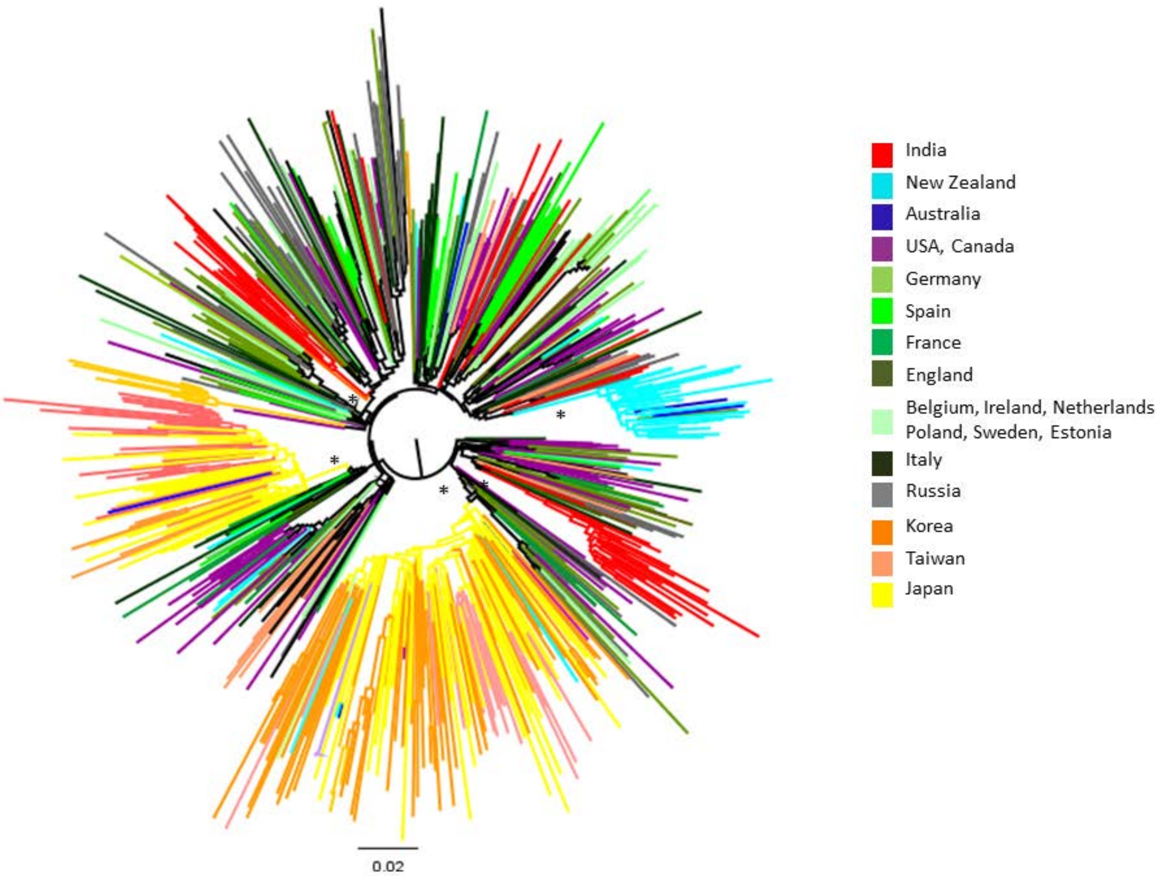

India
New Zealand
Australia
USA, Canada
Germany
Spain
France
England
Belgium, Ireland, Netherlands
Poland, Sweden, Estonia
Italy
Russia
Korea
Taiwan
Japan
0.02

Figure 3.


India
New Zealand
Australia
USA, Canada, Puerto Rico
Germany
Spain
France
England
Netherlands, Poland, Sweden, Estonia
Italy
Russia
0.03


India
New Zealand
Australia
USA, Canada
Germany
Spain
France
England
Belgium, Ireland, Netherlands
Poland, Sweden, Estonia
Italy
Russia
Korea
Taiwan
Japan
0.02